Study for the Carbon Emission Influencing Factors of China Based on Random Forest Model

Jinjin Yang

College of Management and Economics, China University of Petroleum (Beijing)
Email: yangjinjin2022@126.com

ABSTRACT
In order to cope with global climate change and achieve the goal of sustainable development, low-carbon economy has become a new economic development trend. Carbon emissions, as an important indicator of the level of low-carbon economy, are affected by many factors. Therefore, this paper first analyzes the importance of several influencing factors affecting carbon emissions by using random forest regression model, and then proposes to use the number of scientific and technological journals as an indicator to measure the progress of science and technology, and introduces it into the quantitative analysis of the impact factors for characteristic importance analysis. The results show that compared with the growth of forest area and the change in the proportion of energy consumption of new energy sources, the effect of scientific and technological progress on energy conservation and emission reduction is more significant. Based on the analysis of the results, this paper finally provides some inspiration and suggestions for future policy formulation and economic development.

Keywords: *carbon emission, random forest model, scientific and technological progress, low-carbon economy*

1. INTRODUCTION

At present, with the development of the economy and the progress of society, environmental problems are becoming increasingly prominent, especially the global climate change problem with climate warming as the main feature has become one of the problems that need to be solved urgently in the economic development of various countries. As we all know, the main reason for the problem of global warming is the increase in the emission of greenhouse gases such as carbon dioxide, so the development of a low-carbon economy has become an inevitable choice for countries to cope with global climate change and achieve sustainable development. As a developing country, China's economy has developed rapidly since entering the twenty-first century, but its carbon dioxide emissions have also increased significantly. Compared with other industrialized countries, China's carbon emissions have shown a rapid growth trend since 1990. Although in recent years, with the implementation of policies such as energy conservation and emission reduction, the growth rate of Carbon Emissions in China has gradually stabilized. However, due to the impact of China's large economic base, large population and other factors, China's total carbon emissions are still at a high level. Therefore, the low-carbon economy is still the main theme of China's future economic development [1].

Carbon emissions as one of the main indicators to measure the level of low-carbon economic development in various countries has attracted the attention of many scholars, and its relevant research mainly focuses on the analysis of influencing factors, the study of measurement methods, the study of curve fitting and peak prediction[2][3][4] , and the foothold of this paper is the study of influencing factors. Through the search of relevant literature, we found that some scholars advocate the use of qualitative analysis to select and predict the influencing factors of carbon emissions[5]; Some scholars advocate the use of quantitative analysis schemes, based on a variety of indicators, such as carbon emission growth rate, energy consumption change law and carbon emission change trend quantitative analysis, so as to determine the influencing factors, provide policy support, etc. [2][6][7]. In the process of literature review, we found that the application of quantitative analysis in determining influencing factors is more worthy of reference, and at the same time, we also found that although there are many literature considering the important influencing factor of scientific and

Y. Jiang et al. (Eds.): ICEDBC 2022, AEBMR 225, pp. 989–994, 2022.
https://doi.org/10.2991/978-94-6463-036-7_145

technological progress, there is no good measurement index. Therefore, this paper adopts the quantitative analysis method, based on the stochastic forest regression model, first for the annual GDP growth rate, forest area growth rate, and the proportion of clean energy, the three influencing factors related to carbon emissions are analyzed, and then the data source of the number of scientific and technological journals is creatively used as an alternative measure of scientific and technological progress, and the quantitative analysis impact factor is added to the quantitative analysis factor to explore the utility of scientific and technological progress on the reduction of carbon emissions. Finally, based on the analysis results, certain suggestions are made for future policy formulation and economic development.

2. RANDOM FOREST MODEL

The random forest algorithm is an algorithm based on the combination of multiple decision trees, using an integrated thinking method to solve classification problems and regression problems[8], proposed by Leo Breiman and Adele Cutler[9]. It mainly implements its core classification and regression functions through decision trees, each of which is a basic model consisting of a random vector and a training set.

The random forest algorithm process is as follows:

(1) Generate a random vector θ_n for the decision tree which numbered n, that is the basis for the generation of the decision tree;

(2) The original training sample X is sampled by sampling method to construct a sub-training set, and then used the decision tree which numbered n for training, resulting in the sub-set X_n;

(3) The model $h(x,\theta_n)$ for the decision tree which numbered n is generated by using the random vector θ_n and the training set X_n, where x is the independent variable;

(4) After k-round learning, k decision tree vector $h(x,\theta_i), i=1,2,3\cdots k$ is obtained, then this is grouped into a system of multi-classification models by averaging or voting. In this article, the voting method is used by default, and the specific formula is as follows:

$$H(x)=\operatorname{argmax}\sum_{i=1}^{k}h\left(x,\theta_i\right)$$

As a fast and efficient machine learning method, random forest algorithm occupies a very important position in classification and regression applications. Compared with other ensemble learning algorithms, the random forest algorithm has many advantages. Firstly, the accuracy of classification and prediction of random forest algorithm is relatively high; Secondly, the stochastic forest algorithm is not prone to over-fitting, and at the same time, it can focus on the importance of specific features in the process of classification regression, and more comprehensively assess the impact of each feature.

3. EMPIRICAL ANALYSIS

Based on the data of the World Bank, on the basis of removing abnormal years, the data from 2001-2020 are selected for quantitative analysis based on the integrity of the data, and the main factors considered in the paper are GDP growth rate, forest growth rate, and the proportion of new energy in total energy consumption. For these three variables, a random forest with carbon emission change rate is modeled, and the training set and test set are split according to the ratio of 0.15, and then the model prediction and feature importance analysis are carried out.

3.1. Basic random forest model

According to the analysis of the annual GDP growth rate and the annual rate of change of carbon emissions from 2001 to 2020, it can be found from Figure 1 that the correlation between the two is extremely significant, which shows that the growth of China's GDP depends heavily on the consumption of fossil energy. As can be seen from Figure 2, with the economic development, the carbon emissions caused by China's GDP per 100 million yuan are decreasing, which shows that the growth of China's GDP, although it is more dependent on fossil energy in the early stage, but with the development of economic development and scientific and technological progress, in the near future, the quality of GDP has been significantly improved.

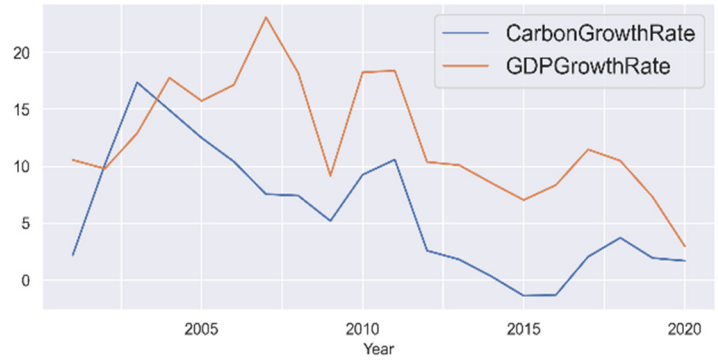

Figure 1 The rate of change in China's GDP and carbon emissions from 1992 to 2015

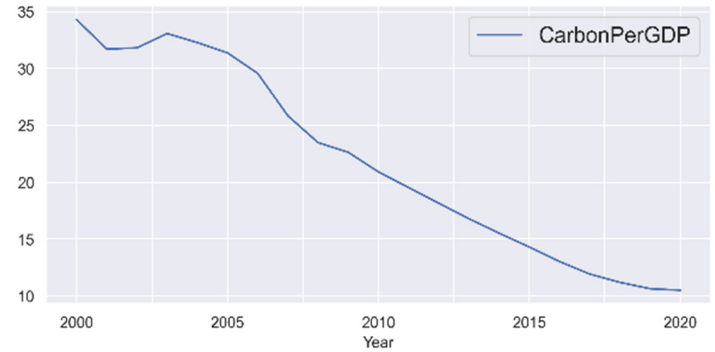

Figure 2 The proportion of carbon emissions consumed per 100 million yuan of GDP in China

When the rate of change in the number of scientific and technical journals is not considered, after random forest regression analysis based on data intervals,the features random forest importance of the growth rate of GDP, forests, new energy is shown in Figure 3. From the figure 3, we can see that the growth rate of forests has the greatest impact on the change of carbon emissions among the three factors, and the rising proportion of new energy in overall energy consumption has also had a significant impact on carbon emissions. At the same time, based on past data, it can be found that the impact of GDP growth rate on carbon emissions is declining, that is, China's low-carbon economy has achieved better results.

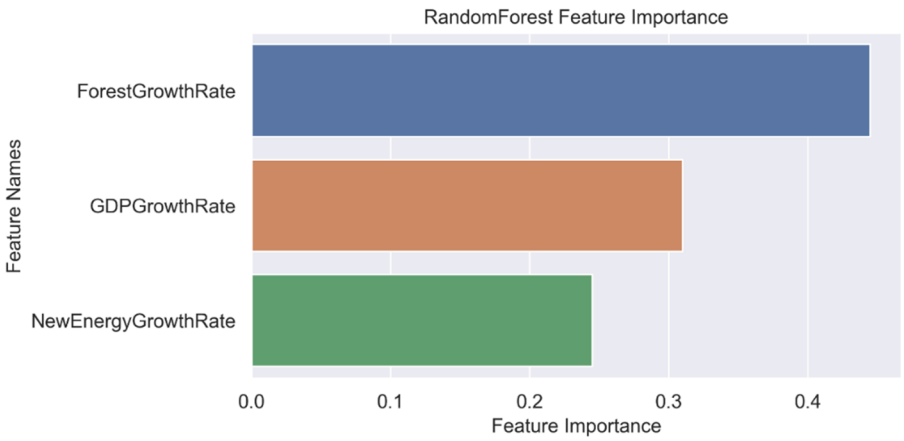

Figure 3 the features random forest importance of the growth rate of GDP, forests, new energy from 2001 to 2020

3.2. A random forest model that introduces technological variable

The data source of the number of scientific and technological journals is used as an alternative measure of scientific and technological progress, and add it to the quantitative analysis impact factor.According to the data from 2001 to 2020, Calculate the growth rate of carbon emissions, GDP, forests, new energy as a proportion and the number of scientific and technological journals , as shown in Table 1.

Table 1. the growth rate of carbon emission, GDP, forest, new energy as a proportion of energy and the number of scientific and technological journals from 2001 to 2020

year	Growth rate of carbon emission (%)	Growth rate of GDP (%)	Growth rate of forest (%)	Growth rate of new energy as a proportion (%)	Growth rate of the number of scientific and technical journals (%)
2001	5.462174	10.55342	1.333885	-3.84003	0.327433
2002	7.984121	9.790723	1.316326	-6.89998	0.050776
2003	15.88544	12.90252	1.299224	-9.10895	0.174233
2004	16.05443	17.76871	1.282561	-6.23403	0.384878
2005	13.61408	15.74315	1.26632	66.44086	0.373647
2006	10.53063	17.147	1.250484	30.22538	0.148872
2007	8.649064	23.08339	1.23504	23.14678	0.132969
2008	2.95927	18.19833	1.219973	80.28628	0.157252
2009	7.228621	9.169499	1.205269	50.34599	0.149864
2010	9.791954	18.24915	1.190916	29.52165	0.091297
2011	9.535132	18.39781	0.965441	27.70781	0.04561
2012	2.747932	10.37827	0.956209	24.32003	0.006869
2013	4.232258	10.09753	0.947152	34.1433	0.091967
2014	-0.42006	8.533392	0.938265	13.79518	0.086625
2015	-0.65195	7.038177	0.929544	19.72915	0.045027
2016	-0.0891	8.3525	1.0298	5.0708	0.07883
2017	-0.0845	11.4739	0.8847	1.2825	0.07718
2018	-0.0612	10.4857	0.8769	3.121	0.0133
2019	-0.05	7.3138	0.8693	5.3544	0.1477
2020	-0.0125	2.9874	0.8618	3.4496	0.99601

Based on the above data, random forest regression modeling is carried out, and it is found that the characteristic variable of the growth rate of the number of scientific and technological journals is added, and its importance far exceeds the rate of change of forest growth. Its importance is as high as 0.47, while the characteristic importance of the forest change rate is 0.22, while the importance of the GDP change rate and the proportion of new energy sources in total energy consumption is 0.17 and 0.13, respectively. This result shows that the impact of scientific and technological progress on carbon emission changes is significant and long-term. The random forest importance of each feature is shown in Figure 4.

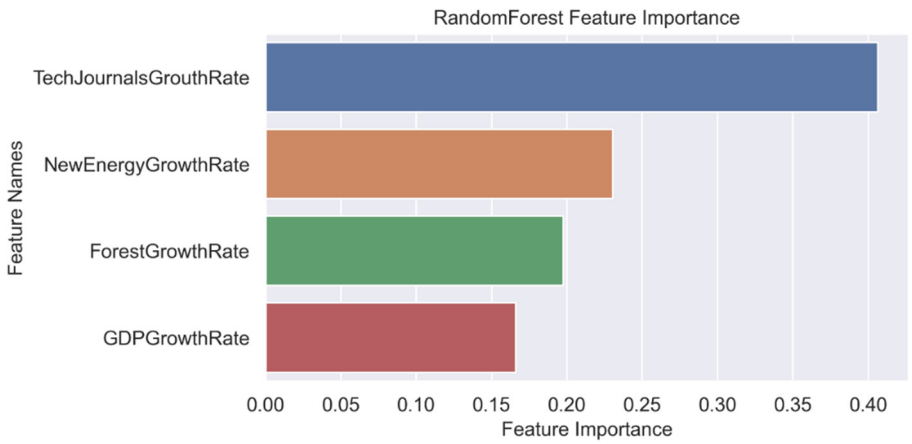

Figure 4 the features random forest importance of the growth rate of GDP, forests, new energy and the number of scientific and technological journal from 2001 to 2020

4. CONCLUSION

In this paper, the significance of the impact factors of random forests on carbon emissions in China is analyzed, and it is found that the number of scientific and technological journals representing scientific and technological progress has a significant impact on the change of carbon emissions. According to the ranking of the characteristic importance of the model, it can also be found that forest change has a more positive impact on carbon emissions, while the use of new energy requires more policy encouragement and support. What can really promote energy conservation and emission reduction is the progress of science and technology. The scientific and technological innovation and scientific revolution triggered by the investment in science and technology can not only maintain rapid economic growth, but also reduce carbon emissions. And the acceleration of economic growth, accompanied by a reduction in carbon emissions, is also a good indication of the improvement in the quality of GDP. So, combined with the results of model analysis, there are the following three points of inspiration for China's future economic development:

(1) Pay attention to the development of science and technology. Science and technology is the primary productive force, give full play to the role of science and technology, not only can maintain the rapid growth of the economy, but also can greatly reduce carbon dioxide emissions, promote the virtuous circle of the environment and economy.

(2) Attach importance to forest protection. Forest greening has a great role in promoting energy conservation and emission reduction, while protecting cultivated land, it is also necessary to actively promote forest protection and urban greening;

(3) Economic development and environmental protection coexist. The control and reduction of carbon emissions is extremely important for the future of the country, and economic growth is important, but if we blindly allow the high-emission and high-pollution resource-consuming economic growth model, we will inevitably fall into the dilemma of economic growth retardation and poor environment.

REFERENCES

[1] Shi H, Chai J, Lu Q, et al. The impact of China's low-carbon transition on economy, society and energy in 2030 based on CO2 emissions drivers[J]. Energy, 2022, 239.

[2] Jiang H. An Analysis of the Influencing Factors of Carbon Dioxide Emissions -- Base on the Consumption Choice[J]. Journal of Physics: Conference Series, 2020, 1578(1):012223 (6pp).

[3] Li B, Shi Y, Hao J, et al. Research on a Carbon Emission Calculation Model and Method for an Underground Fully Mechanized Mining Process[J]. Energies, 2022, 15.

[4] Zhou W, Zeng B, Wang J, et al. Forecasting Chinese carbon emissions using a novel grey rolling prediction model[J]. Chaos, Solitons & Fractals, 2021, 147.

[5] Dai Y. An Empirical Study of Carbon Emission Impact Factors Based on the Vector Autoregression Model[J]. Energies, 2021, 14.

[6] Quan C, Chen X, Yu S, et al. Analysis on the influencing factors of carbon emission in China's logistics industry based on LMDI method - ScienceDirect[J]. Science of The Total Environment, 2020, 734.

[7] Yang W, Gao H, Yang Y. Analysis of Influencing Factors of Embodied Carbon in China's Export Trade in the Background of "Carbon Peak" and "Carbon Neutrality"[J]. Sustainability, 2022, 14.

[8] Zhang Z, Cai Z. Permeability Prediction of Carbonate Rocks Based on Digital Image Analysis and Rock Typing Using Random Forest Algorithm[J]. Energy & Fuels, 2021(3).

[9] Breiman L. Random forests[J]. MACH LEARN, 2001 ,45(1):5-32.